

Equivalence for Networks with Adversarial State

Oliver Kosut, *Member, IEEE* and Jörg Klierer *Senior Member, IEEE*

Abstract

We address the problem of finding the capacity of noisy networks with either independent point-to-point compound channels (CC) or arbitrarily varying channels (AVC). These channels model the presence of a Byzantine adversary which controls a subset of links or nodes in the network. We derive equivalence results showing that these point-to-point channels with state can be replaced by noiseless bit-pipes without changing the network capacity region. Exact equivalence results are found for the CC model, and for some instances of the AVC, including all nonsymmetrizable AVCs. These results show that a feedback path between the output and input of a CC can increase the equivalent capacity, and that if common randomness can be established between the terminals of an AVC (either by feedback, a forward path, or via a third-party node), then again the equivalent capacity can increase. This leads to an observation that deleting an edge of arbitrarily small capacity can cause a significant change in network capacity. We also analyze an example involving an AVC for which no fixed-capacity bit-pipe is equivalent.

I. INTRODUCTION

One fundamental problem in wireless and wireline networks is to achieve robustness against active adversaries. A common assumption is to consider Byzantine adversaries who observe all transmissions, messages, and channel noise values and interfere with the transmitted signals, i.e., by replacing a subset of the channel output values or by injecting additional noise to a specific

O. Kosut is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: okosut@asu.edu).

J. Klierer is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (email: jklierer@njit.edu).

This work was presented in part at the 2014 IEEE International Symposium on Information Theory.

This work was supported in part by the U.S. National Science Foundation under grants CCF-1439465 and CCF-1440014.

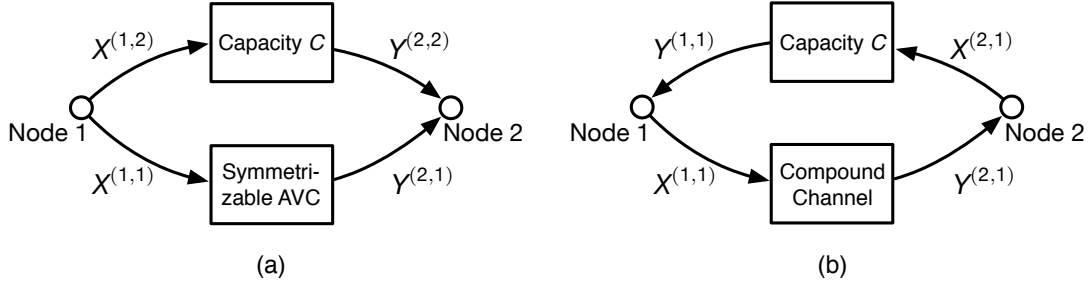


Fig. 1. Two-node networks with a capacity C channel and (a) a symmetrizable AVC, (b) a CC. In general, the upper channel can be replaced with a single-source single-sink network having the same rate.

subset of communication channels or nodes (the adversarial set) in the network. For example, for the adversarial noiseless case both in-network error correction approaches and capacity results under network coding have been presented, e.g., in [1], [2], [3], [4].

The underlying uncertainty in the network due to the action of the adversary leads to channels with varying state in the adversarial set [5]. One possible model is to assume that the corresponding nodes have no knowledge about the exact channel state, but only that the state is selected from a finite set. In the case of a compound channel (CC) [6], [7] the selected state is fixed over the whole transmission of a codeword. In contrast, if the channel state varies from symbol to symbol in an unknown and arbitrary manner we have the case of an arbitrarily varying channel (AVC) [8], [9], [10], [11].

Note that the AVC has a (deterministic) capacity which is either zero or equals the random coding capacity [9]. The former case holds for a symmetrizable AVC, since such a channel can mimic a valid input sequence in such a way that it is impossible for the decoder to decide on the correct codeword. Even though transmission is not possible if such an AVC is considered in isolation, the situation changes in a network setting, as exemplarily depicted in Fig. 1(a). In this two-node network, source and destination nodes are connected via two parallel channels, a (fixed) channel with capacity C and a symmetrizable AVC. Here, communication over the AVC is possible with a non-zero rate since common randomness with negligible rate $\epsilon > 0$ can be shared between both nodes [9], [10], [11] via the upper channel in Fig. 1(a). In a more general setup, in Fig. 1 this channel can be replaced with a single-source single-sink network of positive rate C .

In the following we consider the problem of reliable communication over a network of

independent noisy point-to-point channels in the presence of active adversaries. A subset of the channels either consists of AVCs or CCs. This is in contrast to the model in [12], where the action of the adversary is directly modeled by injecting an arbitrary vector to the network edges in the adversarial set. By building on the results in [13] we identify cases where the adversarial capacity of the network equals the capacity of another network in which each channel is replaced by a noise-free bit-pipe. For a CC, the bit-pipe has capacity equal to the standard CC capacity if there is no feedback path from the output to the input; if there is, then the equivalent bit-pipe has higher capacity, because the state can be estimated at the output and relayed back to the input (see Fig. 1(b)). For an AVC, the equivalent bit-pipe has capacity equal to the random coding capacity if it is possible to establish common randomness between the input and output. This can be accomplished if any of the following hold: (i) the AVC is non-symmetrizable, (ii) there is a parallel forward path as in Fig. 1(a), (iii) there is a feedback path as for the CC in Fig. 1(b), or (iv) a third-party node can transmit to both the input and output nodes. If none of these hold, it appears to be difficult to obtain an equivalence result, as the strong converse does not hold for symmetrizable AVCs. Indeed, we illustrate in Sec. VIII that there exist AVC networks in which no equivalent bit-pipe with fixed capacity exists.

The structure of the paper is as follows. In Sec. II, we formally introduce the problem for both CC and AVC models. In Sec. III we describe the concept of stacked networks, introduced in [13], and state two preliminary lemmas. In Sec. IV, we introduce a lemma demonstrating that training sequences can be used for the CC model to reliably estimate the channel state. In Sec. V, given a channel model and a pair of nodes u and v , we determine whether it is possible to transmit information at any positive rate from u to v . These results will be used in the equivalence results for both state models: for the CC model, to determine whether feedback is possible, and for the AVC model, whether common randomness can be established (*c.f.* Fig. 1). In Sec. VI we present our main equivalence results for the CC model, and in Sec. VII for the AVC model. In Sec. VIII we analyze an example AVC network that we show has no equivalent bit-pipe. In Sec. IX we relate our results to the edge removal problem, which has proved difficult for state-less networks but we prove has a simple solution for both CC and AVC models. We conclude in Sec. X.

II. MODEL

Consider a network of nodes $\mathcal{V} := \{1, \dots, m\}$ with state, given by

$$\mathcal{N} = \left(\prod_{v=1}^m \mathcal{X}^{(v)}, \mathcal{S}, p(\mathbf{y}|\mathbf{x}, s), \prod_{v=1}^m \mathcal{Y}^{(v)} \right). \quad (1)$$

Herein, $\mathcal{X}^{(v)}$ and $\mathcal{Y}^{(v)}$ denote the input and output alphabets of the node v and \mathcal{S} the set of network states, respectively. This network may represent either a CC or an AVC model. These both assume that the state is chosen not randomly but adversarially; in the CC model the adversary chooses a single state $s \in \mathcal{S}$ that remains constant throughout the code block, whereas in the AVC model the adversary chooses an arbitrary state sequence $s^n \in \mathcal{S}^n$. In this paper we are interested in both problems, but only one at a time. Studying networks with both CC-type state and AVC-type state is beyond our scope. Further we assume that the overall set of network states decomposes into a product of the set of states for the link being replaced by a bit pipe and the set of states for the rest of the network. Relaxing this assumption would significantly complicate the problem, as the capacity of a point-to-point channel could be coupled to the behavior of the rest of the network. In such a case it may not be possible to have equivalence between a noisy point-to-point channel and a bit-pipe with fixed capacity.

In general, CCs and AVCs can be quite pathological, so we assume that alphabets $\mathcal{X}^{(v)}$, \mathcal{S} , and $\mathcal{Y}^{(v)}$ are all finite sets. Most of our results apply for more general alphabets under mild regularity conditions, but to avoid edge cases and complications we restrict ourselves to finite alphabets. We believe that the interesting consequences of the CC and AVC network models are captured with finite alphabets models, and that the complications that arise for general alphabets are unlikely to make a difference in practice.

Notation: Let $[k] = \{1, \dots, k\}$. A rate vector \mathcal{R} consists of multicast rates $R^{\{v\} \rightarrow U}$ from each source node v to each destination set $U \subseteq \mathcal{V}$. With a singleton destination set $U = \{u\}$, we sometimes write simply $R^{(v \rightarrow u)}$. For each (v, U) pair, there is a message $W^{\{v\} \rightarrow U} \in \mathcal{W}^{\{v\} \rightarrow U} = [2^{nR^{\{v\} \rightarrow U}}]$. Let $W^{(V \rightarrow *)}$ denote the vector of all messages originating at nodes $v \in V$, and let $\mathcal{W}^{(V \rightarrow *)}$ denote the corresponding message set. Also let W denote the vector of all messages. For a set $\mathcal{A} \subset \mathcal{V}$, we write $X^{(\mathcal{A})} = (X^{(v)} : v \in \mathcal{A})$, and similarly for $Y^{(\mathcal{A})}$. We also write \mathbf{x} for $X^{(\mathcal{V})}$ and \mathbf{y} for $Y^{(\mathcal{V})}$, as in (1).

A blocklength- n solution $S(\mathcal{N})$ for network \mathcal{N} consists of a set of causal encoding functions

$$X_t^{(v)} : (\mathcal{Y}^{(v)})^{t-1} \times \mathcal{W}^{(\{v\} \rightarrow *)} \rightarrow \mathcal{X}^{(v)} \quad (2)$$

for each $v \in \mathcal{V}$ and $t \in [n]$, and decoding functions

$$\widehat{W}^{(\{v\} \rightarrow U), u} : (\mathcal{Y}^{(u)})^n \times \mathcal{W}^{(\{u\} \rightarrow *)} \rightarrow \mathcal{W}^{(\{v\} \rightarrow U)} \cup \{e\} \quad (3)$$

for each (v, U) pair and each $u \in U$, where e is a special symbol that denotes declaring an error. Let \widehat{W} be the complete vector of message estimates, and denote by $\{\widehat{W} \neq W\}$ the event that at least one message is incorrectly decoded. Note that the probability of this event depends on the state sequence S^n .

Definition 1: The CC-capacity region $\mathcal{R}_{\text{CC}}(\mathcal{N})$ of network \mathcal{N} is given by the closure of the set of rate vectors \mathcal{R} for which there exists a sequence of blocklength- n solutions for which

$$\max_{s \in \mathcal{S}} \Pr(\widehat{W} \neq W | S^n = (s, s, \dots, s)) \rightarrow 0. \quad (4)$$

Definition 2: The AVC-capacity region $\mathcal{R}_{\text{AVC}}(\mathcal{N})$ of network \mathcal{N} is given by the closure of the set of rate vectors \mathcal{R} for which there exists a sequence of blocklength- n solutions for which

$$\max_{s^n \in \mathcal{S}^n} \Pr(\widehat{W} \neq W | S^n = s^n) \rightarrow 0. \quad (5)$$

It is easy to see that neither $\mathcal{R}_{\text{CC}}(\mathcal{N})$ nor $\mathcal{R}_{\text{AVC}}(\mathcal{N})$ change if the state is allowed to be randomized instead of deterministic, as long as this random choice is independent of the message and the operation of the channel, and for the CC model the state is fixed across the coding block.

We are especially interested in the case that there is an independent point-to-point channel from node 1 to node 2 with independent state. That is, $\mathcal{X}^{(1)} = \mathcal{X}^{(1,0)} \times \mathcal{X}^{(1,1)}$, $\mathcal{Y}^{(2)} = \mathcal{Y}^{(2,0)} \times \mathcal{Y}^{(2,1)}$, $\mathcal{S} = \mathcal{S}^{(0)} \times \mathcal{S}^{(1)}$, and

$$p(\mathbf{y} | \mathbf{x}, s) = p(\mathbf{y}^{(0)} | \mathbf{x}^{(0)}, s^{(0)}) p(y^{(2,1)} | x^{(1,1)}, s^{(1)}) \quad (6)$$

where $x^{(1,1)} \in \mathcal{X}^{(1,1)}$, $y^{(2,1)} \in \mathcal{Y}^{(2,1)}$, and $s^{(1)} \in \mathcal{S}^{(1)}$ represent the input, output, and state respectively for the point-to-point channel, and $\mathbf{x}^{(0)} \in \mathcal{X}^{(1,0)} \times \prod_{v \neq 1} \mathcal{X}^{(v)}$, $\mathbf{y}^{(0)} \in \mathcal{Y}^{(2,0)} \times \prod_{v \neq 2} \mathcal{Y}^{(v)}$, and $s^{(0)} \in \mathcal{S}^{(0)}$ represent the input, output, and state respectively for the remainder of the network. The point-to-point channel itself is given by

$$\mathcal{C} = (\mathcal{X}^{(1,1)}, \mathcal{S}^{(1)}, p(y^{(2,1)} | x^{(1,1)}, s^{(1)}), \mathcal{Y}^{(2,1)}). \quad (7)$$

We also consider the network \mathcal{N}^R for any $R \geq 0$ in which the noisy point-to-point channel \mathcal{C} is replaced by a rate- R noiseless (and state-less) bit-pipe denoted \mathcal{C}^R . By convention, for non-integer R , with n uses \mathcal{C}^R can transmit $\lfloor nR \rfloor$ bits.

Our goal is to prove achievability-type results of the form $\mathcal{R}(\mathcal{N}^R) \subseteq \mathcal{R}(\mathcal{N})$ and converse-type results of the form $\mathcal{R}(\mathcal{N}) \subseteq \mathcal{R}(\mathcal{N}^R)$ for both CC and AVC models.

III. STACKED NETWORKS

We adopt the notion from [13] of *stacked networks*, wherein we denote by $\underline{\mathcal{N}}$ a network with N independent copies of the network \mathcal{N} . Each copy (layer) contains an instance of every channel input and every channel output, all operating independently¹. Underlines denote stacked variables and vectors, and the argument ℓ refers to layer ℓ , where $\ell \in [N]$. That is, $\underline{X}^{(v)}(\ell)$ is the symbol transmitted by node v in layer ℓ , and $\underline{Y}^{(v)}(\ell)$ is the symbol received by node v in layer ℓ . Moreover, we denote $\underline{X}^{(v)} = (\underline{X}^{(v)}(\ell) : \ell \in [N])$ and similarly for $\underline{Y}^{(v)}$. The corresponding alphabets are given by $\underline{\mathcal{X}}^{(v)}$, etc. Message sets are correspondingly increased by a factor of N ; that is, $\underline{\mathcal{W}}^{\{\{v\} \rightarrow U\}} = (\mathcal{W}^{\{\{v\} \rightarrow U\}})^N$. Rates are therefore defined by $R^{\{\{v\} \rightarrow U\}} = |\underline{\mathcal{W}}^{\{\{v\} \rightarrow U\}}|/(nN)$.

We need to differentiate between the CC and AVC models for stacked networks, because for the CC model the state remains constant across time and across layers, whereas for the AVC model the state may vary between layers. For the CC model, the distribution of channel outputs $\underline{\mathbf{Y}} = (\underline{Y}^{(v)} : v \in \mathcal{V})$ given channel inputs $\underline{\mathbf{X}} = (\underline{X}^{(v)} : v \in \mathcal{V})$ and state $s \in \mathcal{S}$ is

$$p(\underline{\mathbf{y}}|\underline{\mathbf{x}}, s) = \prod_{\ell=1}^N p(\underline{\mathbf{y}}(\ell)|\underline{\mathbf{x}}(\ell), s) \quad (8)$$

where $\underline{\mathbf{X}}(\ell)$ and $\underline{\mathbf{Y}}(\ell)$ are the vectors of transmitted and received symbols respectively in layer ℓ . For the AVC model, there is a different state in each layer denote $\underline{S}(\ell)$ for layer ℓ . The distribution of $\underline{\mathbf{Y}}$ given $\underline{\mathbf{X}}$ and state vector $\underline{S} = (\underline{S}(\ell) : \ell \in [N])$ is

$$p(\underline{\mathbf{y}}|\underline{\mathbf{x}}, \underline{s}) = \prod_{\ell=1}^N p(\underline{\mathbf{y}}(\ell)|\underline{\mathbf{x}}(\ell), \underline{s}(\ell)). \quad (9)$$

Solutions for stacked networks are defined similarly to those for unstacked networks, the only difference being that each coding function has access to all stacks from prior time instances.

¹With the exception that in the CC model, the state is constant across all layers of the network and all time.

In particular, the transmitted symbols for all layers at node v and time t are determined by the causal encoding function

$$\underline{X}_t^{(v)} : (\underline{\mathcal{Y}}^{(v)})^{t-1} \times \underline{\mathcal{W}}^{(\{v\} \rightarrow *)} \rightarrow \underline{\mathcal{X}}^{(v)} \quad (10)$$

and the decoding function for message $\underline{W}^{(\{v\} \rightarrow U)}$ at node $u \in U$ is given by

$$\widehat{\underline{W}}^{(\{v\} \rightarrow U), u} : (\underline{\mathcal{Y}}^{(u)})^n \times \underline{\mathcal{W}}^{(\{u\} \rightarrow *)} \rightarrow \underline{\mathcal{W}}^{(\{u\} \rightarrow U)} \cup \{e\}. \quad (11)$$

Note that node v has access to its received symbols and messages in all layers when deciding its transmissions. The capacity regions for the stacked networks $\mathcal{R}_{\text{CC}}(\underline{\mathcal{N}})$ and $\mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}})$ are defined analogously as above for unstacked networks.

The following two preliminary lemmas are simple extensions of Lemmas 1 and 4 respectively from [13] to include state.

Lemma 1: For any network \mathcal{N} , $\mathcal{R}_{\text{CC}}(\mathcal{N}) = \mathcal{R}_{\text{CC}}(\underline{\mathcal{N}})$ and $\mathcal{R}_{\text{AVC}}(\mathcal{N}) = \mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}})$.

Proof sketch: Following along the proof of Lemma 1 in [13], $\mathcal{R}_{\text{CC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{CC}}(\underline{\mathcal{N}})$ and $\mathcal{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}})$ can be proved by repeating any solution on network \mathcal{N} over all N layers in the stacked network $\underline{\mathcal{N}}$. Since any solution for \mathcal{N} must work for any (CC-type or AVC-type) state, the repeated solution on $\underline{\mathcal{N}}$ will also work. To prove $\mathcal{R}_{\text{CC}}(\mathcal{N}) \supseteq \mathcal{R}_{\text{CC}}(\underline{\mathcal{N}})$ and $\mathcal{R}_{\text{AVC}}(\mathcal{N}) \supseteq \mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}})$, we may take any solution on $\underline{\mathcal{N}}$ and implement it on \mathcal{N} with blocklength nN by correctly “unraveling” the solution. Again, the effects of the unknown state are the same in each. ■

Lemma 2: The capacity regions $\mathcal{R}_{\text{CC}}(\mathcal{N}^R)$ and $\mathcal{R}_{\text{AVC}}(\mathcal{N}^R)$ are continuous in R for all $R > 0$.

Proof: We employ a very similar proof technique as that of Lemma 4 in [13]. By Lemma 1, it is equivalent to prove continuity for $\mathcal{R}_{\text{CC}}(\underline{\mathcal{N}}^R)$ and $\mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}}^R)$. Fix any $\delta \in (0, R)$ and rate vector $\mathcal{R} \in \text{int}(\mathcal{R}_{\text{CC}}(\underline{\mathcal{N}}^{R+\delta}))$ (resp. $\mathcal{R} \in \text{int}(\mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}}^{R+\delta}))$). Assume that $\underline{\mathcal{N}}^{R+\delta}$ has N layers. Let $\underline{\mathcal{N}}^{R-\delta}$ be an N' -fold stacked network with

$$N'(R - \delta) \geq N(R + \delta). \quad (12)$$

For all $\lambda > 0$, there exists solution $S(\underline{\mathcal{N}}^{R+\delta})$ with rate vector \mathcal{R} and probability of error λ . We define a solution $S(\underline{\mathcal{N}}^{R-\delta})$ based on $S(\underline{\mathcal{N}}^{R+\delta})$ as follows. Use precisely the same coding operations aside from the bit-pipe $\underline{\mathcal{C}}^{R+\delta}$ for the first N layers of the stack, and send the $\lfloor N(R+\delta) \rfloor$ bits to be sent across $\underline{\mathcal{C}}^{R+\delta}$ instead across the bit-pipe $\underline{\mathcal{C}}^{R-\delta}$. This can be done because of (12).

Note that the resulting rate vector for $\mathcal{S}(\underline{\mathcal{N}}^{R-\delta})$ is

$$\mathcal{R}' = \frac{\mathcal{R}N}{N'} > \mathcal{R} \frac{N}{N(R+\delta)/(R-\delta)+1}. \quad (13)$$

Thus the difference between \mathcal{R} and \mathcal{R}' vanishes as $N \rightarrow \infty$ and $\delta \rightarrow 0$.

Recall that for the CC-model (resp. AVC-model), the state does not affect operation of the bit-pipes. Meanwhile, as the rest of the network is operated identically in the two solutions—aside from the $N' - N$ unused layers in the solution on $\underline{\mathcal{N}}^{R-\delta}$ —the effect of the state is precisely the same. Thus the modified solution on $\underline{\mathcal{N}}^{R-\delta}$ has precisely the same probability of error λ . Therefore $\mathcal{R}' \in \mathcal{R}_{\text{CC}}(\underline{\mathcal{N}}^{R-\delta})$ (resp. $\mathcal{R}' \in \mathcal{R}_{\text{AVC}}(\underline{\mathcal{N}}^{R-\delta})$). ■

IV. COMPOUND CHANNEL TRAINING LEMMA

The following lemma will be used several times in CC results. It asserts that CC states can be estimated using training sequences.

Lemma 3: Fix a point-to-point CC $(\mathcal{X}, \mathcal{S}, p(y|x, s), \mathcal{Y})$. Let $\alpha_{1:n}$ be a training sequence drawn randomly and uniformly from \mathcal{X}^n . Let $Y_{1:n}$ be the output of the channel with input $\alpha_{1:n}$ and state s . Define the following subsets of \mathcal{S} :

$$\hat{\mathcal{S}} = \arg \max_{s' \in \mathcal{S}} p(Y_{1:n} | \alpha_{1:n}, s'), \quad (14)$$

$$\bar{\mathcal{S}} = \{s' \in \mathcal{S} : p(y|x, s') = p(y|x, s) \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}\}. \quad (15)$$

Then

$$\lim_{n \rightarrow \infty} \Pr(\hat{\mathcal{S}} \neq \bar{\mathcal{S}}) = 0. \quad (16)$$

Proof: We may write

$$\hat{\mathcal{S}} = \arg \min_{s' \in \mathcal{S}} -\frac{1}{n} \sum_{t=1}^n \log p(Y_t | \alpha_t, s'). \quad (17)$$

Given true state s , for any $s' \in \mathcal{S}$, the quantities $-\log p(Y_t | \alpha_t, s')$ are i.i.d. with expected value

$$\sum_{x,y} -p(y|x, s) \log p(y|x, s') = H(Y|X, S=s) + \Delta_{s,s'} \quad (18)$$

where

$$\Delta_{s,s'} := \sum_{x \in \mathcal{X}} \frac{1}{|\mathcal{X}|} D(p(y|x, s) \| p(y|x, s')). \quad (19)$$

Note that $\Delta_{s,s'} = 0$ if and only if $s' \in \bar{\mathcal{S}}$. Let $\delta_s = \min\{\Delta_{s,s'} : \Delta_{s,s'} > 0\}$. We have $\delta_s > 0$ since \mathcal{S} is finite. Let $\tau_s = H(Y|X, S = s) + \delta_s/2$. Define the events

$$\mathcal{E}_1 := \left\{ -\frac{1}{n} \sum_{t=1}^n \log p(Y_t | \alpha_t, s') \geq \tau_s \text{ for any } s' \in \bar{\mathcal{S}} \right\} \quad (20)$$

$$\mathcal{E}_2 := \left\{ -\frac{1}{n} \sum_{t=1}^n \log p(Y_t | \alpha_t, s') \leq \tau_s \text{ for any } s' \notin \bar{\mathcal{S}} \right\}. \quad (21)$$

Recall that $\Delta_{s,s'} > 0$ implies $\Delta_{s,s'} \geq \delta_s$. The event in (16) implies $\mathcal{E}_1 \cup \mathcal{E}_2$, so it is enough to show they each have probability vanishing in n . This follows from the Law of Large Numbers and the fact that \mathcal{S} is finite. \blacksquare

V. POSITIVE RATE REGIONS

For both CC and AVC models, it will be important to know whether any information at all can be sent between nodes. This positive (but arbitrarily small) rate will be used for feedback in the CC model and generating shared randomness in the AVC model (see Fig. 1). Thus in this section we investigate the set of node pairs (u, v) for which positive rate can be sent from u to v . We do this first without state, and then extend it for the CC and AVC models.

A. Positive Rate Without State

Assume for now that \mathcal{S} contains only a single element, in which case $\mathcal{R}_{\text{CC}}(\mathcal{N}) = \mathcal{R}_{\text{AVC}}(\mathcal{N})$, and we denote both by $\mathcal{R}(\mathcal{N})$. We form a set $\mathcal{P} \subset \mathcal{V} \times \mathcal{V}$ and subsequently show that \mathcal{P} is precisely the set of node pairs that can sustain positive rate. For the CC model, we will be interested in whether $(2, 1) \in \mathcal{P}$; *i.e.* whether feedback is possible with respect to the point-to-point channel from node 1 to node 2. On the other hand, for the AVC model, we care whether there exists a node u such that $(u, 1), (u, 2) \in \mathcal{P}$.

The set \mathcal{P} is formed via the following steps:

- 1) Initialize \mathcal{P} as $\{(u, u) : u \in \mathcal{V}\}$.
- 2) If there is a pair of nodes $(u, v) \notin \mathcal{P}$, and a set $\mathcal{A} \subset \mathcal{V}$ such that $(j, v) \in \mathcal{P}$ for all $j \in \mathcal{A}$, and

$$\max_{p(x^{(u)}), x^{\{\mathcal{A}\}^c}} I(X^{(u)}; Y^{(\mathcal{A})} | X^{\{\mathcal{A}\}^c} = x^{\{\mathcal{A}\}^c}) > 0, \quad (22)$$

then add (u, v) to \mathcal{P} .

3) Repeat step 2 until there are no additional such pairs (u, v) .

The mutual information in (22) represents the capacity of a point-to-point channel with input $X^{(u)}$ and output $Y^{(\mathcal{A})}$, even though $Y^{(\mathcal{A})}$ represents all received values by nodes in \mathcal{A} , which are not available at any single receiver. Additionally, we maximize over constants $x^{\{\{u\}^c\}}$ in case the channel from $X^{(u)}$ to $Y^{(\mathcal{A})}$ only has positive capacity for certain transmissions by the other nodes.

Theorem 4: If $(u, v) \in \mathcal{P}$, then there exists an $\mathcal{R} \in \mathcal{R}(\mathcal{N})$ with $R^{(u \rightarrow v)} > 0$.

Proof: A detailed proof is given by the proof of the stronger result Lemma 7, to be stated below. Roughly, the solution is derived as follows. A node may trivially send arbitrary amounts of information to itself; thus $R^{(u \rightarrow u)} > 0$ is achievable for any $u \in \mathcal{V}$. We proceed by induction to prove the theorem for pairs $(u, v) \in \mathcal{P}$ with $u \neq v$. Consider the specific step in the construction of \mathcal{P} at which (u, v) is added, and let \mathcal{A} satisfy (22). We assume that for all $j \in \mathcal{A}$, positive rate can be sent from j to v . To send positive rate from u to v , we employ a point-to-point channel code from $X^{(u)}$ to $Y^{(\mathcal{A})}$. A message is chosen at node u , and the corresponding codeword is transmitted by node u and received by nodes in \mathcal{A} . Next, the received sequences are transmitted from nodes in \mathcal{A} to node v using positive-rate solutions that are assumed to exist by the induction hypothesis and since by construction $(j, v) \in \mathcal{P}$ for all $j \in \mathcal{A}$. Finally, node v decodes the point-to-point code. ■

The following theorem gives the converse result, stating that if $(u, v) \notin \mathcal{P}$, then values received at node v are conditionally independent of values sent from node u given messages that originate outside node u . This indicates that all information known at node v originates outside of node u ; i.e., the input at node u cannot influence the output at node v . This is a much stronger statement than a simple converse, and indeed even stronger than a usual “strong” converse, but it is necessary to prove equivalence results.

Theorem 5: If $(u, v) \notin \mathcal{P}$, then for any solution $\mathcal{S}(\mathcal{N})$, $X_{1:n}^{(u)} \rightarrow W^{\{\{u\}^c \rightarrow *\}} \rightarrow Y_{1:n}^{(v)}$ forms a Markov chain.

Proof: Fix $(u, v) \notin \mathcal{P}$. Let $\mathcal{A} := \{i : (i, v) \in \mathcal{P}\}$. By the definition of \mathcal{P} , for any $i \notin \mathcal{A}$,

$$\max_{p(x^{\{\{i\}\}}, x^{\{\{i\}^c\}})} I(X^{\{\{i\}\}}; Y^{(\mathcal{A})} | X^{\{\{i\}^c\}} = x^{\{\{i\}^c\}}) = 0. \quad (23)$$

As this holds for all $i \notin \mathcal{A}$, we conclude that for any solution $\mathcal{S}(\mathcal{N})$, we have the Markov chain

$$X_t^{(\mathcal{A}^c)} \rightarrow X_t^{(\mathcal{A})} \rightarrow Y_t^{(\mathcal{A})} \quad (24)$$

for each time t . We may now write

$$p\left(y_{1:n}^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}\right) = \prod_{t=1}^n p\left(y_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) \quad (25)$$

$$= \prod_{t=1}^n \sum_{x_t^{(\mathcal{V})}} p\left(x_t^{(\mathcal{V})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{V})}\right) \quad (26)$$

$$= \prod_{t=1}^n \sum_{x_t^{(\mathcal{V})}} p\left(x_t^{(\mathcal{V})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{A})}\right) \quad (27)$$

$$= \prod_{t=1}^n \sum_{x_t^{(\mathcal{A})}} p\left(x_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{A})}\right) \quad (28)$$

$$= \prod_{t=1}^n \sum_{x_t^{(\mathcal{A})}} p\left(x_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{A})}\right) \quad (29)$$

$$= \prod_{t=1}^n p\left(y_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, y_{1:t-1}^{(\mathcal{A})}\right) = p\left(y_{1:n}^{(\mathcal{A})} \middle| w^{(\mathcal{A})}\right) \quad (30)$$

where (27) follows from (24), and (29) follows by the dependency requirements of the coding at nodes in \mathcal{A} . From this derivation, we conclude that $X_{1:n}^{(\mathcal{A}^c)} \rightarrow W^{(\mathcal{A})} \rightarrow Y_{1:n}^{(\mathcal{A})}$ forms a Markov chain. This completes the proof since $v \in \mathcal{A}$ and $u \in \mathcal{A}^c$. \blacksquare

Theorems 4 and 5 completely determine when any positive rate is achievable, as stated in the following corollary.

Corollary 6: There exists a rate vector $\mathcal{R} \in \mathcal{R}(\mathcal{N})$ with $R^{(\{v\} \rightarrow U)} > 0$ if and only if $(v, i) \in \mathcal{P}$ for all $i \in U$.

Note that the “only if” direction of Corollary 6 is weaker than Theorem 5, because even if $R^{(v \rightarrow u)}$ cannot be positive, it does not mean that the strong statement of Theorem 5 holds.

B. Positive Rate for the CC Model

We now extend the above results for CC-type state. For each $s \in \mathcal{S}$, define \mathcal{P}_s as above for \mathcal{P} , but with fixed state $S = s$. Let $\mathcal{P}_{\text{CC}} = \bigcap_{s \in \mathcal{S}} \mathcal{P}_s$.

For any state s such that $(u, v) \in \mathcal{P}_s$, the following lemma establishes the existence of solutions for the CC model with positive rate from u to v such that (i) if the state is s , node v can reliably decode the message; and (ii) if the state is *not* s , node v either decodes correctly or declares an error. Recall that we use the symbol e to signify a decoder declaring an error. We construct

these solutions using training sequences (*c.f.* Lemma 3), wherein node v only decodes if s is among the most likely states. Thus if the true state is not s , either node v will discover this and declare an error, or the channel is indistinguishable from that with state s , so node v will decode reliably. The solutions from this lemma will be used to prove that positive rate can be transmitted from u to v for $(u, v) \in \mathcal{P}_{\text{CC}}$.

Lemma 7: For any state $s \in \mathcal{S}$, and all $(u, v) \in \mathcal{P}_s$, there exist a sequence of solutions $S_{u,v,s}^{(n)}(\mathcal{N})$ with rate $R^{(u \rightarrow v)} > 0$ such that

- 1) if $S = s$ then the probability of error vanishes with n , and
- 2) if $S \neq s$ then the probability of making an error without declaring an error (*i.e.* that $\widehat{W}^{(u \rightarrow v)} \notin \{W^{(u \rightarrow v)}, e\}$) vanishes with n .

Proof: We adopt the convention that a node may send arbitrary amounts of information to itself; thus the lemma is immediate if $u = v$. We proceed by induction to prove the theorem for pairs $(u, v) \in \mathcal{P}_s$ with $u \neq v$. Consider the specific step in the construction of \mathcal{P}_s at which (u, v) was added. There is a set $\mathcal{A} \subset \mathcal{V}$ such that for some distribution $p(x^{(u)})$ and constant $x^{\{\{u\}^c\}}$,

$$I(X^{(u)}; Y^{(\mathcal{A})} | X^{\{\{u\}^c\}} = x^{\{\{u\}^c\}}, S = s) > 0 \quad (31)$$

and (j, v) for all $j \in \mathcal{A}$ has already been added to \mathcal{P}_s . We assume there exist sequences of solutions $S_{j,v,s}^{(n_j)}(\mathcal{N})$ for all $j \in \mathcal{A}$, with rates $R^{(j \rightarrow v)} > 0$, satisfying the probability of error constraints in the statement of the lemma. Fix a length n to be determined later.

We now describe the coding procedure. Initially node u chooses a message $W^{(u \rightarrow v)} \in \mathcal{W}^{(u \rightarrow v)} = [2^{n\tilde{R}^{(u \rightarrow v)}}]$, where $\tilde{R}^{(u \rightarrow v)}$ is any positive number strictly smaller than the mutual information in (31). Coding proceeds in 3 sessions, described as follows. The lengths of the first two sessions are n , and that of the third session is $\sum_{j \in \mathcal{A}} n_j$. Thus the quantity $\tilde{R}^{(u \rightarrow v)}$ is not the rate achieved by the code, because the overall blocklength is longer than n .

Session 1: Node u transmits a training sequence $\alpha_{1:n}$ drawn randomly and uniformly from $(\mathcal{X}^{(u)})^n$ while other nodes transmit the constant $x^{\{\{u\}^c\}}$. The training sequence constitutes part of the codebook and is revealed to all nodes prior to coding. For each $j \in \mathcal{A}$, let $Y_{1:n}^{(j)}$ be the received sequence at node j for each $j \in \mathcal{A}$.

Session 2: Node u transmits $W^{(u \rightarrow v)}$ via an n -length point-to-point channel code from $X^{(u)}$ to $Y^{(\mathcal{A})}$ with input distribution $p(x^{(u)})$ and distribution conditioned on $X^{\{\{u\}^c\}} = x^{\{\{u\}^c\}}$ and

$S = s$, while all other nodes transmit the constant $x^{\{u\}^c}$. Let $Y_{n+1:2n}^{(j)}$ be the received sequence at node j at each $j \in \mathcal{A}$.

Session 3: Dividing into $|\mathcal{A}|$ sub-sessions, we run one sub-session for each $j \in \mathcal{A}$, in which $S_{j,v,s}^{(n_j)}(\mathcal{N})$ is employed to transmit $Y_{1:2n}^{(j)}$ from j to v , where the blocklength is given by

$$n_j = \left\lceil \frac{2n \log |\mathcal{Y}^{(j)}|}{R^{j \rightarrow v}} \right\rceil \quad (32)$$

so that $2^{n_j R^{(j \rightarrow v)}} \geq |\mathcal{Y}^{(j)}|^{2n}$. Let $\hat{Y}_{1:2n}^{(j)}$ be the decoded sequence at node v .

Decoding: If any of the solutions $S_{j,v,s}^{(n_j)}(\mathcal{N})$ declares an error, then node v declares an error. Otherwise, given $\hat{Y}_{1:n}^{(\mathcal{A})}$ node v determines whether s is among the most likely states given the training sequence; that is

$$s \in \arg \max_{s'} p(\hat{Y}_{1:n}^{(\mathcal{A})} | \alpha_{1:n}, X^{\{u\}^c} = x^{\{u\}^c}, S = s'). \quad (33)$$

If (33) does not hold, then node v declares an error. If it does, then node v decodes the message from $\hat{Y}_{n+1:2n}^{(\mathcal{A})}$ using the point-to-point channel decoder. Let $\widehat{W}^{(u \rightarrow v)}$ be the decoded message.

Achieved rate: Recall that all we need to show is that the achieved rate $R^{(u \rightarrow v)}$ is positive. The total blocklength for the code is $2n + \sum_{j \in \mathcal{A}} n_j$, so the overall rate is given by

$$R^{(u \rightarrow v)} = \frac{n \tilde{R}^{(u \rightarrow v)}}{2n + \sum_{j \in \mathcal{A}} n_j} \quad (34)$$

$$\geq \frac{\tilde{R}^{(u \rightarrow v)}}{2 + \sum_{j \in \mathcal{A}} \frac{2 \log |\mathcal{Y}^{(j)}|}{R^{j \rightarrow v}} + \frac{|\mathcal{A}|}{n}}. \quad (35)$$

Note that $R^{(u \rightarrow v)}$ is bounded above 0 for sufficiently large n .

Probability of error analysis: First consider the case that $S = s$. We need to show that $\Pr(\widehat{W}^{(u \rightarrow v)} \neq W^{(u \rightarrow v)})$ can be made arbitrarily small. Define the error events

$$\mathcal{E}_1 := \left\{ \hat{Y}_{1:2n}^{(\mathcal{A})} \neq Y_{1:2n}^{(\mathcal{A})} \right\} \quad (36)$$

$$\mathcal{E}_2 := \left\{ s \notin \arg \max_{s'} p(\hat{Y}_{1:n}^{(\mathcal{A})} | \alpha_{1:n}, X^{\{u\}^c} = x^{\{u\}^c}, S = s') \right\} \quad (37)$$

$$\mathcal{E}_3 := \left\{ \widehat{W}^{(u \rightarrow v)} \neq W^{(u \rightarrow v)} \right\}. \quad (38)$$

The overall error event is \mathcal{E}_3 , and we can upper bound its probability by

$$\Pr(\mathcal{E}_3 | S = s) \leq \Pr(\mathcal{E}_1 | S = s) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2 | S = s) + \Pr(\mathcal{E}_3 | \mathcal{E}_1^c, \mathcal{E}_2^c, S = s). \quad (39)$$

By the inductive assumptions that $S_{j,v,s}^{(n_j)}(\mathcal{N})$ have vanishing probability of error given state s for each $j \in \mathcal{A}$, $\Pr(\mathcal{E}_1) \rightarrow 0$. By Lemma 3, $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2 | S = s) \rightarrow 0$. Finally, $\Pr(\mathcal{E}_3 | \mathcal{E}_1^c, \mathcal{E}_2^c, S = s)$ is merely the probability of error of the point to point code from u to \mathcal{A} , so it vanishes as $n \rightarrow \infty$. Thus the overall probability of error may be made arbitrarily small.

Now consider the case that $S = \bar{s} \neq s$. Define the additional error events

$$\mathcal{E}_4 := \left\{ \text{solution } S_{j,v,s}^{(n_j)}(\mathcal{N}) \text{ declares an error for some } j \in \mathcal{A} \right\} \quad (40)$$

$$\mathcal{E}_5 := \left\{ \widehat{W}^{(u \rightarrow v)} \notin \{W^{(u \rightarrow v)}, e\} \right\}. \quad (41)$$

We need to show $\Pr(\mathcal{E}_5) \rightarrow 0$ as $n \rightarrow \infty$. If either \mathcal{E}_2 or \mathcal{E}_4 occurs, then node v declares an error, so $\mathcal{E}_5 \subset \mathcal{E}_2^c \cap \mathcal{E}_4^c$. In addition, $\mathcal{E}_5 \subset \mathcal{E}_3$, so

$$\Pr(\mathcal{E}_5 | S = \bar{s}) \leq \Pr(\mathcal{E}_3 \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c | S = \bar{s}) \quad (42)$$

$$\leq \Pr(\mathcal{E}_1 \cap \mathcal{E}_4^c | S = \bar{s}) + \Pr(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c | S = \bar{s}). \quad (43)$$

The first term in (43) vanishes by the inductive assumption on $S_{j,v,s}^{(n_j)}(\mathcal{N})$ for all $j \in \mathcal{A}$. To bound the second term, we consider two cases. First, that $p(y^{(\mathcal{A})} | x^{(u)}, x^{\{\mathcal{A}\}^c}, \bar{s}) \neq p(y^{(\mathcal{A})} | x^{(u)}, x^{\{\mathcal{A}\}^c}, s)$ for any $x^{(u)} \in \mathcal{X}^{(u)}$ and $y^{(\mathcal{A})} \in \mathcal{Y}^{(\mathcal{A})}$. Then $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c | S = \bar{s}) \rightarrow 0$ by Lemma 3. Otherwise, the channel from $x^{(u)}$ to $Y^{(\mathcal{A})}$ conditioned on $X^{\{\mathcal{A}\}^c} = x^{\{\mathcal{A}\}^c}$ is identical for $S = \bar{s}$ and $S = s$. Hence the operation of the point-to-point code from $X^{(u)}$ to $Y^{(\mathcal{A})}$ works just as well for $S = \bar{s}$ as for $S = s$, so $\Pr(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c | S = \bar{s}) \rightarrow 0$. ■

The following theorem gives the positive rate result (equivalent to Theorems 4 and 5) for the CC model.

Theorem 8: If $(u, v) \in \mathcal{P}_{\text{CC}}$, then there exists a rate vector $\mathcal{R} \in \mathcal{R}_{\text{CC}}(\mathcal{N})$ with $R^{(u \rightarrow v)} > 0$. Conversely, if $(u, v) \notin \mathcal{P}_{\text{CC}}$, then for any solution $\mathcal{S}(\mathcal{N})$ there exists $s \in \mathcal{S}$ such that with $S^n = (s, s, \dots, s)$, $X_{1:n}^{(u)} \rightarrow W^{\{\mathcal{A}\}^c \rightarrow *} \rightarrow Y_{1:n}^{(v)}$ forms a Markov chain.

Proof: To prove the converse, note that if $(u, v) \notin \mathcal{P}_{\text{CC}}$ then $(u, v) \notin \mathcal{P}_s$ for some $s \in \mathcal{S}$. With this fixed state, the proof follows exactly as that of Theorem 5.

Now we prove achievability. Suppose $(u, v) \in \mathcal{P}_{\text{CC}}$. Thus $(u, v) \in \mathcal{P}_s$ for all $s \in \mathcal{S}$. Let $S_{u,v,s}^{(n)}(\mathcal{N})$ be the sequence of solutions asserted by Lemma 7. Let $R_s^{(u \rightarrow v)} > 0$ be the rate for code $S_{u,v,s}^{(n)}(\mathcal{N})$. Let $\tilde{R}^{(u \rightarrow v)} = \min_{s \in \mathcal{S}} R_s^{(u \rightarrow v)}$.

We construct a solution to send positive rate from u to v as follows. First node u chooses a message $W^{(u \rightarrow v)} \in [2^{n\tilde{R}^{(u \rightarrow v)}}]$. Coding proceeds in $|\mathcal{S}|$ sessions. In the session associated with

$s \in \mathcal{S}$, we employ $S_{u,v,s}^{(n)}(\mathcal{N})$ to send $W^{(u \rightarrow v)}$ from u to v . After all sessions are complete, node v decodes by choosing $\widehat{W}^{(u \rightarrow v)}$ to be the output of the first solution that did not declare an error. By Lemma 7, with high probability the solution associated with the true state will not make an error, and any solution associated with a false state will not make an error without declaring an error. Thus the probability of error is small. As the total blocklength for the code is $n|\mathcal{S}|$, the achieved rate is $\tilde{R}^{(u \rightarrow v)}/|\mathcal{S}| > 0$. ■

C. Positive Rate for the AVC Model

Recall that, as defined in [10], an AVC $p(y|x, s)$ is *symmetrizable* if there exists a probability transition matrix $p(s|x)$ such that

$$\sum_{s \in \mathcal{S}} p(y|x, s)p(s|x') = \sum_{s \in \mathcal{S}} p(y|x', s)p(s|x), \text{ for all } x, x' \in \mathcal{X}, y \in \mathcal{Y}. \quad (44)$$

As shown in [10], a point-to-point AVC has positive capacity if and only if it is non-symmetrizable. Now define \mathcal{P}_{AVC} using the same procedure as above for \mathcal{P} , but replace (22) with the condition that there exists $x^{\{\{u\}^c\}} \in \mathcal{X}^{\{\{u\}^c\}}$ such that the channel from $X^{(u)}$ to $Y^{(\mathcal{A})}$, conditioned on $X^{\{\{u\}^c\}} = x^{\{\{u\}^c\}}$, is non-symmetrizable.

Theorem 9: If $(u, v) \in \mathcal{P}_{\text{AVC}}$, then there exists a rate vector $\mathcal{R} \in \mathcal{R}_{\text{AVC}}(\mathcal{N})$ with $R^{(u \rightarrow v)} > 0$.

Proof: The proof follows from the same argument as for Theorem 4, except that we replace the point-to-point channel code from $X^{(u)}$ to $Y^{(\mathcal{A})}$ with an AVC code. By the assumption that this channel is non-symmetrizable, positive rate can be achieved by the results in [10]. ■

VI. COMPOUND CHANNEL EQUIVALENCE

In this section and the next we simplify notation by writing X for $X^{(1,1)}$, Y for $Y^{(2,1)}$, and S for $S^{(1)}$. Since we are primarily interested in the independent channel \mathcal{C} , there should be no confusion.

There are two relevant capacities for the compound channel: first, the standard capacity expression for a compound channel

$$\underline{C} = \max_{p(x)} \min_{s \in \mathcal{S}} I(X; Y|S = s), \quad (45)$$

and second, the capacity of a compound channel if the state is known at the encoder and the decoder, wherein the min and max are reversed:

$$\bar{C} = \min_{s \in \mathcal{S}} \max_{p(x)} I(X; Y|S = s). \quad (46)$$

Of course, $\underline{C} \leq \bar{C}$. Let \mathcal{P}_{CC} be defined as above for \mathcal{N} . As stated in the following theorem, the compound channel is equivalent to a bit-pipe with rate either \underline{C} or \bar{C} , depending on whether the rest of the network can sustain any positive feedback rate from node 2 to node 1.

Theorem 10:

$$\mathcal{R}_{\text{CC}}(\mathcal{N}) = \begin{cases} \mathcal{R}_{\text{CC}}(\mathcal{N}^{\bar{C}}) & \text{if } (2, 1) \in \mathcal{P}_{\text{CC}} \\ \mathcal{R}_{\text{CC}}(\mathcal{N}^{\underline{C}}) & \text{if } (2, 1) \notin \mathcal{P}_{\text{CC}}. \end{cases} \quad (47)$$

We prove this theorem in several lemmas, which in combination with continuity from Lemma 2 prove the theorem.

Lemma 11: If $R < \underline{C}$, then $\mathcal{R}_{\text{CC}}(\mathcal{N}^R) \subseteq \mathcal{R}_{\text{CC}}(\mathcal{N})$.

Proof: The proof follows an almost identical argument as that of Lemma 5 from [13], which proved that a bit-pipe may simulate a point-to-point noisy channel via a traditional channel code. Recalling that \underline{C} is the usual compound channel capacity, $R < \underline{C}$ implies the existence of a reliable compound channel code at rate R . Replacing the channel code in the proof of Lemma 5 from [13] with such a compound channel code proves our result. ■

Lemma 12: If $R > \bar{C}$, then $\mathcal{R}_{\text{CC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{CC}}(\mathcal{N}^R)$.

Proof: Let $s^* = \arg \min_s \max_{p(x)} I(X; Y)$. We may use Theorem 6 in [13], which proves that a bit-pipe can simulate a noisy channel with less capacity, to simulate the channel $p(y|x, s^*)$ over the bit-pipe of rate R , since $R > I(X; Y)$ for this channel and any input distribution. ■

Lemma 13: If $(2, 1) \in \mathcal{P}_{\text{CC}}$ and $R < \bar{C}$, then $\mathcal{R}_{\text{CC}}(\mathcal{N}^R) \subseteq \mathcal{R}_{\text{CC}}(\mathcal{N})$.

Proof: By Theorem 4, since $(2, 1) \in \mathcal{P}_{\text{CC}}$, there exists a solution $S_0(\mathcal{N})$ such that $R^{(2 \rightarrow 1)} > 0$. Given a solution $S(\mathcal{N}^R)$, we construct a solution $S(\mathcal{N})$ with three sessions. In session 1, node 1 sends a training sequence so that node 2 can learn the state. In session 2, this estimated state is transmitted back to node 1 using $S_0(\mathcal{N})$. In session 3, node 1 uses this estimated state to transmit a message across \mathcal{C} while the rest of $S(\mathcal{N}^R)$ is conducted. We give more details as follows.

Session 1: We employ a random coding argument wherein we choose a training sequence $\alpha_{1:n_1}$ randomly and uniformly from \mathcal{X}^{n_1} . This sequence forms the codebook for session 1, and it is revealed to nodes 1 and 2. Node 1 transmits $\alpha_{1:n_1}$ into \mathcal{C} while the inputs to all other channels are arbitrary. Let $Y_{1:n_1}$ be the output of \mathcal{C} . Node 2 forms a state estimate by choosing \hat{S} arbitrarily from the set $\hat{S} := \arg \max_{s' \in \mathcal{S}} p(Y_{1:n_1} | \alpha_{1:n_1}, s')$.

Session 2: Employ $S_0(\mathcal{N})$ with blocklength n_2 to transmit \hat{S} from node 2 to node 1. Let \check{S} be the recovered value at node 1. Assume n_2 is large enough such that $2^{n_2 R^{(2 \rightarrow 1)}} \geq |\mathcal{S}|$.

Session 3: While the rest of the network conducts $S(\mathcal{N}^R)$, node 1 employs an encoder for the point-to-point channel with state \check{S} , while node 2 employs a decoder for the channel with state \hat{S} . Let n_3 be the blocklength of this session.

Probability of error analysis: Assume the state is s . Define the following error events:

$$\mathcal{E}_1 := \{p(y|x, s) \neq p(y|x, \hat{S}) \text{ for any } x, y\} \quad (48)$$

$$\mathcal{E}_2 := \{\check{S} \neq \hat{S}\} \quad (49)$$

$$\mathcal{E}_3 := \{\widehat{W}^{(1 \rightarrow 2)} \neq W^{(1 \rightarrow 2)}\}. \quad (50)$$

We may bound the probability of error by

$$\Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3). \quad (51)$$

By Lemma 3, $\Pr(\mathcal{E}_1) \rightarrow 0$ as $n_1 \rightarrow \infty$. By Theorem 4, $\Pr(\mathcal{E}_2) \rightarrow 0$ as $n_2 \rightarrow \infty$. The effective rate of the point-to-point code in Session 3 is $\frac{nR}{n_3}$, where the total blocklength is $n = n_1 + n_2 + n_3$. Since by assumption $R < \bar{C}$, for sufficiently large $n_3/(n_1 + n_2)$ the effective rate is bounded below \bar{C} . Moreover, $\bar{C} \leq \max_{p(x)} I(X; Y|S = s)$, so the effective rate is bounded below the capacity of the point-to-point channel with state s . As long as \mathcal{E}_1 and \mathcal{E}_2 do not hold, then $\hat{S} = \check{S}$ are a state for which the operation of the channel is identical to that of s , so the channel with this state has the same capacity as with s . Hence $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3) \rightarrow 0$ as $n_3 \rightarrow \infty$. ■

The following theorem is essentially equivalent to Theorem 4 in [13], but with a compound channel instead of a standard channel without state.

Lemma 14: If $(2, 1) \notin \mathcal{P}_{\text{CC}}$ and $R > \underline{C}$, then $\mathcal{R}_{\text{CC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{CC}}(\mathcal{N}^R)$.

Proof: By Lemma 1 it suffices to show that $\mathcal{R}_{\text{CC}}(\underline{\mathcal{N}}) \subseteq \mathcal{R}_{\text{CC}}(\underline{\mathcal{N}}_R)$. Fix any $\mathcal{R} \in \text{int}(\mathcal{R}_{\text{CC}}(\mathcal{N}))$ and $\lambda > 0$.

Choose code and define distributions: Let $S(\mathcal{N})$ be a rate- \mathcal{R} solution on network \mathcal{N} for some blocklength n . By Theorem 8, for solution $S(\mathcal{N})$, $X_{1:n}^{(2)} \rightarrow W^{\{\{2\}^c \rightarrow *\}} \rightarrow Y_{1:n}^{(1)}$ forms a Markov chain. Moreover, the state S only has direct impact on $Y_{1:n}^{(2)}$, which in turn only has direct impact on $X_{1:n}^{(2)}$. Thus $S \rightarrow X_{1:n}^{(2)} \rightarrow (W^{\{\{2\}^c \rightarrow *\}}, Y_{1:n}^{(1)})$ forms a Markov chain.² Combining these two

²We have written S as a random variable even though it is arbitrary rather than random. By $S \rightarrow A \rightarrow B$ we mean that $p(b|a, s) = p(b|a)$.

chains yields

$$S \rightarrow X_{1:n}^{(2)} \rightarrow W^{(\{2\}^c \rightarrow *)} \rightarrow Y_{1:n}^{(1)}. \quad (52)$$

Since $W^{(\{2\}^c \rightarrow *)}$ is drawn uniformly from $\mathcal{W}^{(\{2\}^c \rightarrow *)}$ and independently from S , the distribution of $(W^{(\{2\}^c \rightarrow *)}, Y_{1:n}^{(1)})$ does not depend on S . Thus the distribution of $X_{1:n}^{(1)}$ also does not depend on S , as it is a function of $(W^{(\{1\} \rightarrow *)}, Y_{1:n}^{(1)})$. Therefore, for each time t we may define $p_t(x)$ to be the distribution of $X_t^{(1)}$ independent of S . Let $p(x) = \frac{1}{n} \sum_{t=1}^n p_t(x)$ and let

$$s^* = \arg \min_{s \in \mathcal{S}} I(X; Y | S = s) \quad (53)$$

where X is drawn from $p(x)$. Let $p_t(x, y) = p_t(x)p(y|x, s^*)$.

Typical set: Define $\hat{A}_{\epsilon, t}^{(N)}$ to be the N -length typical set according to distribution $p_t(x, y)$ as in [13].

Design of channel emulators: By concavity of mutual information with respect to the input variable,

$$\frac{1}{n} \sum_{t=1}^n I(X_t; Y_t | S = s^*) \leq I(X; Y | S = s^*) = \min_s I(X; Y | S = s) \leq \underline{C} < R. \quad (54)$$

Let $R_t := I(X_t; Y_t | S = s^*) + \Delta$ where $\Delta > 0$ is chosen so that $\frac{1}{n} \sum_{t=1}^n R_t = R$.

Randomly design decoder $\beta_{N,t} : [2^{NR_t}] \rightarrow \mathcal{Y}$ by drawing codewords $\beta_{N,t}(1), \dots, \beta_{N,t}(2^{NR_t})$ from the i.i.d. distribution with marginal $p_t(y)$. Define encoder $\alpha_{N,t} : \mathcal{X} \rightarrow [2^{NR_t}]$ as

$$\alpha_{N,t}(\underline{x}) = \begin{cases} k & \text{if } (\underline{x}, \beta_{N,t}(k)) \in \hat{A}_{\epsilon, t}^{(N)} \\ 1 & \text{if } \nexists k \text{ s.t. } (\underline{x}, \beta_{N,t}(k)) \in \hat{A}_{\epsilon, t}^{(N)}. \end{cases} \quad (55)$$

Note that the number of bits required to send $(\alpha_{N,t}(\underline{X}))_{t=1}^n$ is $\sum_{t=1}^n NR_t = nNR$, so we may send all these encoded functions via a bit-pipe of rate R .

The rest of the proof follows essentially that of Theorem 6 in [13]. This involves creating a stacked solution for $\underline{\mathcal{N}}$ with exponentially decreasing probability of error, and then converting it into a solution for $\underline{\mathcal{N}}^R$ by employing the channel emulators at nodes 1 and 2 to simulate the noisy channel over the rate- R bit pipe. Finally, the error probability can be bounded provided correct parameters are chosen for the typical set $\hat{A}_{\epsilon, t}^{(N)}$, which can be done for our problem by virtue of the fact that $R_t - I(X_t; Y_t | S = s^*) = \Delta > 0$. ■

VII. ARBITRARILY VARYING CHANNEL EQUIVALENCE

The random coding capacity of a point-to-point AVC is defined as the maximum rate that can be achieved if the encoder and decoder have access to shared randomness (inaccessible to the adversary). It is given by

$$C_r = \max_{p(x)} \min_{p(s)} I(X; Y). \quad (56)$$

Moreover, the max and min may be interchanged without changing the quantity, because of the convexity properties of the mutual information. Without shared randomness, as shown in [10], the capacity of an AVC is 0 if the channel is symmetrizable, and C_r if not. Thus, in all cases, C_r is an upper bound on the capacity. The following theorem provides the corresponding network-level converse.

Theorem 15: $\mathcal{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.

Proof: By the continuity property from Lemma 2, it will be enough to show that $\mathcal{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{AVC}}(\mathcal{N}^R)$ for all $R > C_r$. Let

$$p^*(s) := \arg \min_{p(s)} \max_{p(x)} I(X; Y). \quad (57)$$

Let $p^*(y|x) = \sum_s p^*(s)p(y|x, s)$. Note that C_r is the capacity of the ordinary channel with transition matrix $p^*(y|x)$. Let $\mathcal{S}(\mathcal{N})$ be any solution for the original network \mathcal{N} , with probability of error (maximized over state sequences s^n) P_e . Thus $\mathcal{S}(\mathcal{N})$ achieves probability of error P_e for any random choice of s^n , provided this random choice is independent of the choice of message. In particular, the probability of error is no larger than P_e if S^n is drawn i.i.d. from $p^*(s)$. Thus, the capacity region can only enlarge if the AVC is replaced by the ordinary channel $p^*(y|x)$ in \mathcal{N} . Now the proof is completed by Theorem 6 of [13]. ■

Theorem 12.11 from [11] states that the capacity of a point-to-point AVC is either 0 or C_r . This is shown by proving that a small header can be transmitted from encoder to decoder that allows the encoder and decoder to simulate common randomness. This small header can be sent using any code that achieves positive rate. The following is an extension of this result to the network setting wherein the header may originate at any node and be transmitted to both nodes 1 and 2.

Theorem 16: If for some node u , there exists a rate vector $\mathcal{R}_1 \in \mathcal{R}_{\text{AVC}}(\mathcal{N})$ with $R_1^{(u \rightarrow 1)} > 0$ and a rate vector $\mathcal{R}_2 \in \mathcal{R}_{\text{AVC}}(\mathcal{N})$ with $R_2^{(u \rightarrow 2)} > 0$, then $\mathcal{R}_{\text{AVC}}(\mathcal{N}) = \mathcal{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.

Before proving the theorem, we state the following lemma, asserting that C_r can be achieved with a random code that requires a relatively small amount of shared randomness between encoder and decoder. This lemma is a simple combination of Lemmas 12.8 and 12.10 from [11].

Lemma 17: Given an AVC $p(y|x, s)$, for any $R < C_r$ and $\epsilon > 0$, for any integer K satisfying

$$K \geq \frac{2n}{\epsilon}(R + \log |\mathcal{S}|). \quad (58)$$

there exist K different $(n, 2^{nR})$ channel codes (f_ℓ, ϕ_ℓ) for $\ell = 1, \dots, K$ consisting of functions $f_\ell : [2^{nR}] \rightarrow \mathcal{X}^n$ and $\phi_\ell : \mathcal{Y}^n \rightarrow [2^{nR}]$ such that

$$\max_{m \in [2^{nR}]} \max_{s^n \in \mathcal{S}^n} \frac{1}{K} \sum_{\ell=1}^K \Pr(\phi_\ell(Y^n) \neq m | X^n = f_\ell(m), S^n = s^n) \leq \epsilon. \quad (59)$$

Proof of Theorem 16: In light of Theorem 15, we have only to prove that $\mathcal{R}(\mathcal{N}^R) \subseteq \mathcal{R}(\mathcal{N})$ for all $R < C_r$. The proof of this follows essentially from the same argument as the proof of Theorem 12.11 from [11]. Fix $\epsilon > 0$ and $R < C_r$. By Lemma 17 there exists K channel codes (f_ℓ, ϕ_ℓ) satisfying (59), where $\frac{1}{n} \log K \rightarrow 0$ as $n \rightarrow \infty$.

Let $S_1(\mathcal{N})$ be a solution with $R^{(u \rightarrow 1)} > 0$ and let $S_2(\mathcal{N})$ be a solution with $R^{(u \rightarrow 2)} > 0$. Coding proceeds in three sessions. Initially, node u randomly and uniformly chooses an integer $L \in [K]$. In the first session node u transmits L to node 1 using $S_1(\mathcal{N})$. In the second session node u transmits L to node 2 using $S_2(\mathcal{N})$. Let \hat{L}_1 and \hat{L}_2 be the decoded integer at nodes 1 and 2 respectively. For sufficiently large blocklength the probability $\Pr(\hat{L}_1 \neq L \text{ or } \hat{L}_2 \neq L)$ can be made arbitrarily small. In the third session node 1 using encoder $f_{\hat{L}_1}$ and node 2 uses decoder $g_{\hat{L}_2}$ to transmit over the point-to-point AVC. Since the selection of the channel code is random, by (59), on average the probability of error for the third session is bounded by ϵ . ■

The following corollary provides a sufficient condition for equivalence for the AVC. It follows immediately from Theorem 9 and Theorem 16.

Corollary 18: If there exists node u such that $(u, 1) \in \mathcal{P}_{\text{AVC}}$ and $(u, 2) \in \mathcal{P}_{\text{AVC}}$, then $\mathcal{R}_{\text{AVC}}(\mathcal{N}) = \mathcal{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.

VIII. AVC EXAMPLE NETWORK

This section examines the example network shown in Fig. 2. This network illustrates that when a point-to-point AVC does not satisfy the condition of Corollary 18, it is not necessarily

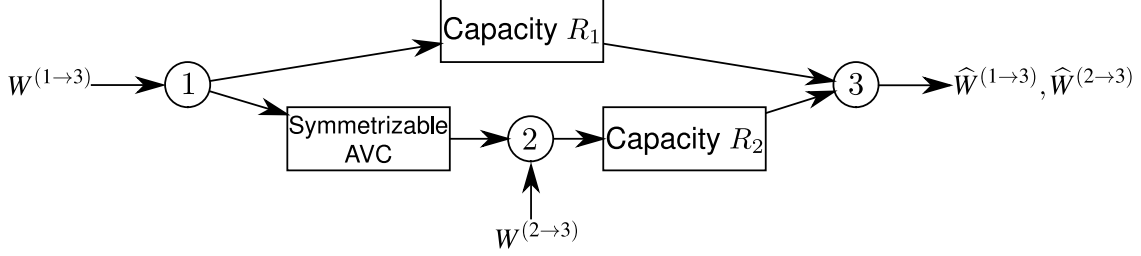


Fig. 2. Example network with a symmetrizable AVC from node 1 to node 2 that does not satisfy the conditions of Corollary 18. The network also contains a rate R_1 bit-pipe between nodes 1 and 3 and a rate R_2 bit-pipe between nodes 2 and 3. Proposition 19 gives the complete capacity region for this network, which cannot be equated to the capacity region of a network in which the AVC is replaced by any bit-pipe of fixed capacity.

equivalent to a zero-capacity bit pipe, or indeed any bit pipe with fixed capacity. The channel from node 1 to node 2 is a symmetrizable AVC given by $p(y|x, s)$, with random code capacity C_r . The channel from node 1 to node 3 is a bit-pipe with capacity R_1 , where we assume $R_1 > 0$, and that from node 2 to node 3 is a bit-pipe with capacity R_2 . We first determine the capacity region of this network, and then find the capacity region if the AVC were replaced by a bit-pipe of capacity fixed capacity \tilde{R} ; these two regions do not coincide for any \tilde{R} . Roughly, equivalence cannot hold because the symmetrizable AVC leads to a situation in which node 2 can determine that the data sent by node 1 is one of a small number of possibilities. All of these possibilities can be sent along link $(2, 3)$, where node 3 can determine which is the correct one using side information from link $(1, 3)$. Thus, as long as R_2 is not too large, each bit sent on link $(2, 3)$ for message $W^{(1 \rightarrow 3)}$ contributes only a fraction of a bit of useful data; no such phenomenon can occur with a fixed-capacity bit-pipe, since an additional bit would add either a full bit or zero bits to the overall capacity.

It was shown in [14] that with list decoding—even for quite short lists—the capacity of a symmetrizable AVC is given by its random code capacity. In particular, [14] defines the *symmetrizability* of an AVC $p(y|x, s)$ as the largest integer M for which there exists a stochastic matrix $p(s|x_1, \dots, x_M)$ such that

$$\sum_{s \in \mathcal{S}} p(y|x, s) p(s|x_1, \dots, x_M) \quad (60)$$

is symmetric in x, x_1, \dots, x_M . A channel is symmetrizable, in the sense formulated in [10] and discussed above in (44), if and only if $M \geq 1$. It is shown in [14] that for an AVC with

symmetrizability M , the decoder can reliably list-decode at rate C_r with list size $M + 1$. This result will be instrumental in our examination of the example network.

For the network shown in Fig. 2, the only positive achievable rates for this network are $R^{(1 \rightarrow 3)}$ and $R^{(2 \rightarrow 3)}$. The following proposition characterizes the capacity region for this network.

Proposition 19: The capacity region for the network shown in Fig. 2 is given by the pairs $(R^{(1 \rightarrow 3)}, R^{(2 \rightarrow 3)})$ satisfying

$$R^{(2 \rightarrow 3)} \leq R_2 \quad (61)$$

$$R^{(1 \rightarrow 3)} \leq R_1 + C_r \quad (62)$$

$$R^{(2 \rightarrow 3)} + (M + 1)R^{(1 \rightarrow 3)} \leq (M + 1)R_1 + R_2. \quad (63)$$

Proof: Achievability: The basic idea of our achievability proof is as follows: node 2 makes use of the list decoding scheme from [14], and then transmits along link $(2, 3)$ the entire list of $M + 1$ potential messages, in addition to message $W^{2 \rightarrow 3}$. Along link $(1, 3)$, we send part of message $W^{1 \rightarrow 3}$, in addition to a small hash that allows node 3 to determine which of the $M + 1$ messages is the true one. That this is possible with a hash of negligible rate is not quite proved by either the results of [14] or Lemma 17, since in neither scenario is there a list decoding followed by a determination of the true message via side information. Here we use a random linear hash to achieve essentially the same effect as the random choice of channel codes in the proof of Lemma 17 [11], but in the context of a list code, as we will show in the following.

Fix rates $R^{(1 \rightarrow 3)}, R^{(2 \rightarrow 3)}$ satisfying (61)–(63) but with strict inequalities. Fix an integer q and a blocklength n . Let \mathbb{F}_{2^q} be the finite field of order 2^q . We express $W^{(1 \rightarrow 3)}$ as a vector of elements of \mathbb{F}_{2^q} as follows. Let $\tilde{R}^{(1 \rightarrow 3)}$ be the largest multiple of $\frac{q}{n}$ no larger than $R^{(1 \rightarrow 3)}$. Clearly $\tilde{R}^{(1 \rightarrow 3)} \geq R^{(1 \rightarrow 3)} - \frac{q}{n}$. Define integers

$$K_1 = \left\lfloor \frac{nR_1}{q} \right\rfloor - 1 \quad (64)$$

$$K_2 = \frac{n\tilde{R}^{(1 \rightarrow 3)}}{q} - K_1. \quad (65)$$

By the assumption that $R_1 > 0$, for n sufficiently large we have $K_1 \geq 1$. Message $W^{(1 \rightarrow 3)}$ is chosen from $[2^{n\tilde{R}^{(1 \rightarrow 3)}}]$ and message $W^{(2 \rightarrow 3)}$ from $[2^{nR^{(2 \rightarrow 3)}}]$. We may denote $W^{(1 \rightarrow 3)} = (W_1, \dots, W_{K_1+K_2})$ where $W_j \in \mathbb{F}_{2^q}$ for all $j \in [K_1 + K_2]$. Note that the W_j are independent and each drawn uniformly from \mathbb{F}_{2^q} . For convenience, we write $W^{K_2} = (W_{K_1+1}, \dots, W_{K_1+K_2})$.

At the start of encoding, node 1 generates a hash of the vector W^{K_2} . The symbol W_1 is used as the random seed for the hash, and the hash itself is given by

$$h = \sum_{j=1}^{K_2} (W_1)^{j-1} W_{K_1+j}. \quad (66)$$

where $(W_1)^{j-1}$ represents exponentiation in the field \mathbb{F}_{2^q} . Encoding and decoding proceeds as follows:

- 1) (h, W_1, \dots, W_{K_1}) is transmitted along link $(1, 3)$.
- 2) W^{K_2} is encoded using an $(M + 1)$ -list code from [14] and the resulting codeword is transmitted into the AVC $(1, 2)$.
- 3) After receiving the output sequence from the AVC, node 2 decodes the $(M + 1)$ -length list, denoted $\widehat{W}_i^{K_2} = (\widehat{W}_{i,K_1+1}, \dots, \widehat{W}_{i,K_1+K_2})$ for $i \in [M + 1]$.
- 4) $(W^{2 \rightarrow 3}, \widehat{W}_i^{K_2} : i \in [M + 1])$ is transmitted across link $(2, 3)$.
- 5) Node 3 receives the vectors transmitted on links $(1, 3)$ and $(2, 3)$ without error. It decodes $W^{2 \rightarrow 3}$ from its received vector on link $(2, 3)$. Given $\widehat{W}_i^{K_2}$ for each $i \in [M + 1]$ received on link $(2, 3)$, node 3 computes

$$\hat{h}_i = \sum_{j=1}^{K_2} (W_1)^{j-1} \widehat{W}_{i,K_1+j}. \quad (67)$$

For the smallest i for which $\hat{h}_i = h$, node 3 declares

$$\widehat{W}^{(2 \rightarrow 3)} = (W_1, \dots, W_{K_1}, \widehat{W}_{i,K_1+1}, \dots, \widehat{W}_{i,K_1+K_2}). \quad (68)$$

where h and W_1, \dots, W_{K_1} were received on link $(1, 3)$.

Bit-pipe capacity limits: We first confirm that in the coding procedure described above, the vectors sent along links $(1, 3)$ and $(2, 3)$ do not exceed the capacities of these bit-pipes. The number of bits sent along link $(1, 3)$ is $(K_1 + 1)q \leq nR_1$, so its capacity constraint is satisfied. The number of bits sent along link $(2, 3)$ is

$$(M + 1)K_2q + nR^{(2 \rightarrow 3)} \leq n(M + 1)R^{(1 \rightarrow 3)} - n(M + 1)R_1 + 2q + nR^{(2 \rightarrow 3)}, \quad (69)$$

where we have used (70). Since (63) holds with a strict inequality, this quantity is at most nR_2 for sufficiently large n .

Probability of error: There are two potential sources of error: (i) the decoded list from the AVC at node 2 does not include the true intended message, and (ii) there exists $i \in [M + 1]$

such that $\hat{h}_i = h$ even though $\widehat{W}_i^{K_2} \neq W^{K_2}$. For the first source of error, note that the number of bits in W^{K_2} is $K_2 q$, so the rate of the list code on the AVC is

$$\frac{K_2 q}{n} = \tilde{R}^{(1 \rightarrow 3)} - \frac{K_1 q}{n} \leq \tilde{R}^{(1 \rightarrow 3)} - R_1 + \frac{2q}{n} \leq R^{(1 \rightarrow 3)} - R_1 + \frac{2q}{n} \quad (70)$$

where in the first inequality we have used the fact that $K_1 \geq \frac{n R_1}{q} - 2$. Since (62) holds with a strict inequality, the quantity in (70) is less than C_r for sufficiently large n . Thus, by the results in [14], the probability that the decoded list does not include the true message vanishes with n .

Now consider the second source of error. The content of the decoded list depends only on W^{K_2} , the state sequence S^n , and the random operation of the AVC. In particular, the list is independent of W_1 . Thus, for any $w^{K_2}, \hat{w}^{K_2} \in \mathbb{F}_{2^q}^{K_2}$

$$\Pr(\hat{h}_i = h | W^{K_2} = w^{K_2}, \widehat{W}_i^{K_2} = \hat{w}^{K_2}) = \Pr\left(\sum_{j=1}^{K_2} (W_1)^{j-1} (\hat{w}_j - w_j) = 0\right). \quad (71)$$

If $w^{K_2} \neq \hat{w}^{K_2}$, then the polynomial in W_1 inside the probability is a nonzero polynomial of degree at most $K_2 - 1$, so it has at most $K_2 - 1$ roots. Since W_{K_2+1} is chosen uniformly from \mathbb{F}_{2^q} , if $w^{K_2} \neq \hat{w}^{K_2}$ then

$$\Pr(\hat{h}_i = h | W^{K_2} = w^{K_2}, \widehat{W}_i^{K_2} = \hat{w}^{K_2}) \leq \frac{K_2 - 1}{2^q}. \quad (72)$$

Therefore, the probability that $\hat{h}_i = h$ for any i satisfying $\widehat{W}_i^{K_2} \neq W^{K_2}$ is at most

$$\frac{(K_2 - 1)M}{2^q}. \quad (73)$$

This can be made arbitrarily small for sufficiently large q .

Converse: Let $(R^{(1 \rightarrow 3)}, R^{(2 \rightarrow 3)})$ be an achievable rate pair. Thus there exists a sequence of solutions $S_n(\mathcal{N})$ of length n , rates $R^{(1 \rightarrow 3)}, R^{(2 \rightarrow 3)}$ and probability of error going to 0 as $n \rightarrow \infty$. In this argument, we use the fact that the capacity region does not change if the state S^n is chosen randomly, as long as this random choice is independent of the message (but it may depend on the code). We consider two specific distributions for S^n under $S_n(\mathcal{N})$ for some n . First, that S^n is chosen randomly from the i.i.d. distribution with marginal $p^*(s)$ defined in (57) as the saddle-point in the random coding capacity. With this choice, the AVC behaves as a (stateless) stationary memoryless channel with transition probability

$$p^*(y|x) = \sum_s p^*(s) p(y|x, s). \quad (74)$$

Note that the channel $p^*(y|x)$ has capacity C_r . Simple applications of the cutset bound yield (61) and (62).

To prove (63), we consider a different distribution on the state. Let $p_{X^n}(x^n)$ be the distribution of the input sequence to the AVC (1, 2) under solution $S_n(\mathcal{N})$. Note that this distribution depends only on the code at node 1, so it is independent of the state S^n of the AVC. The state sequence S^n is drawn from the distribution

$$\sum_{x_1^n, \dots, x_M^n} p_{X^n}(x_1^n) \cdots p_{X^n}(x_M^n) \prod_{i=1}^n p(s_i | x_{1i}, \dots, x_{mi}) \quad (75)$$

where the distribution $p(s|x_1, \dots, x_m)$ is one for which (60) is symmetric. Let $V(y|x, x_1, \dots, x_M)$ be this symmetric distribution. We may assume the existence of virtual variables X_1^n, \dots, X_M^n wherein $X^n, X_1^n, \dots, X_M^n, Y^n$ are distributed according to

$$p_{X^n}(x^n) p_{X^n}(x_1^n) \cdots p_{X^n}(x_M^n) \prod_{i=1}^n V(y_i | x_i, x_{1i}, \dots, x_{mi}). \quad (76)$$

For convenience we write $X_0^n = X^n$. Let $Z_1 \in [2^{nR_1}]$ and $Z_2 \in [2^{nR_2}]$ be the random variables representing the values sent on links (1, 2) and (2, 3) respectively. By Fano's inequality and the data processing inequality,

$$nR^{(2 \rightarrow 3)} \leq I(W^{(2 \rightarrow 3)}; Z_2) + n\epsilon_n \quad (77)$$

where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Applying Fano's inequality again, we have

$$nR^{(1 \rightarrow 3)} = H(W^{(1 \rightarrow 3)}) \quad (78)$$

$$\leq I(W^{(1 \rightarrow 3)}; Z_1, Z_2) + n\epsilon_n \quad (79)$$

$$= I(W^{(1 \rightarrow 3)}; Z_2) + I(W^{(1 \rightarrow 3)}; Z_1 | Z_2) + n\epsilon_n \quad (80)$$

$$\leq I(W^{(1 \rightarrow 3)}; Z_2) + nR_1 + n\epsilon_n \quad (81)$$

$$\leq I(X_0^n; Z_2) + nR_1 + n\epsilon_n \quad (82)$$

where in (82) we have used the fact that $W^{(1 \rightarrow 3)} \rightarrow X_0^n \rightarrow Z_2$ is a Markov chain. By symmetry of (X_0^n, \dots, X_k^n) , we may generalize (82) to write that for all $k \in \{0, \dots, M\}$

$$nR^{(1 \rightarrow 3)} \leq I(X_k^n; Z_2) + nR_1 + n\epsilon_n. \quad (83)$$

We may sum (77) and (83) to find

$$nR^{(2 \rightarrow 3)} + (M+1)nR^{(1 \rightarrow 3)} \quad (84)$$

$$\leq I(W^{(2 \rightarrow 3)}; Z_2) + \sum_{k=0}^M I(X_k^n; Z_2) + (M+1)nR_1 + (M+2)n\epsilon_n \quad (85)$$

$$\leq I(W^{(2 \rightarrow 3)}; Z_2) + \sum_{k=0}^M I(X_k^n; Z_2 | W^{(2 \rightarrow 3)}, X_0^n, \dots, X_{k-1}^n) + (M+1)nR_1 + (M+2)n\epsilon_n \quad (86)$$

$$= I(W^{(2 \rightarrow 3)}, X_0^n, \dots, X_M^n; Z_2) + (M+1)nR_1 + (M+2)n\epsilon_n \quad (87)$$

$$\leq nR_2 + (M+1)nR_1 + (M+2)n\epsilon_n \quad (88)$$

where in (86) we have used the fact that $(W^{(2 \rightarrow 3)}, X_0^n, \dots, X_M^n)$ are mutually independent. Dividing by n and taking the limit as $n \rightarrow \infty$ yields (63). ■

Suppose that in the example network the AVC were replaced by a bit-pipe of capacity \tilde{R} . It is easy to see that the resulting set of achievable $(R^{(1 \rightarrow 3)}, R^{(2 \rightarrow 3)})$ pairs is given by

$$R^{(2 \rightarrow 3)} \leq R_2 \quad (89)$$

$$R^{(1 \rightarrow 3)} \leq R_1 + \tilde{R} \quad (90)$$

$$R^{(1 \rightarrow 3)} + R^{(2 \rightarrow 3)} \leq R_1 + R_2. \quad (91)$$

This region does not correspond to (61)–(62) for any value of \tilde{R} , as long as $M \geq 1$ (i.e., the AVC is symmetrizable). Therefore, the AVC in Fig. 2 is not equivalent to any fixed capacity bit-pipe.

IX. RELATION TO THE “EDGE REMOVAL” PROBLEM

Consider two networks \mathcal{N} and \mathcal{N}' with identical topologies except for a single edge, which has capacity C_e in network \mathcal{N} , but capacity $C'_e = C_e - \delta$ in network \mathcal{N}' . Herein, $\delta > 0$ is a small constant. Particular attention has been devoted recently to the so called *edge removal* problem which describes the special case of this scenario for $C_e = \delta$. It has been shown in [15], [16] that for a variety of demand types for which the network coding capacity can be described by the cut-set bound, the capacity of every cut is reduced by at most δ for each dimension. This means that if a rate vector \mathcal{R} is achievable in network \mathcal{N} , a rate vector $\mathcal{R} - \delta\mathcal{I}$ is achievable in \mathcal{N}' , where \mathcal{I} denotes the unit rate vector. Examples include single and multisource multicast

and single source cases with non-overlapping demands, but also scenarios for which the cut-set bound is not tight, for example a specific class of multiple unicast networks [16]. Further, in [17] the edge removal problem has also been connected to the problem whether a network coding instance allows a reconstruction with ϵ and zero error, respectively. However, so far only various special cases have been considered, and it is not clear how to formulate the edge removal problem for general demands and topologies.

In the following, based on the discussion in Sections VI and VII, we extend the edge removal problem to networks with state. We formulate our result for both the CC and the AVC case in the following theorem.

Theorem 20: Given a network \mathcal{N} with state according to (1) and assume that a non-zero rate vector $\mathcal{R}(\mathcal{N})$ is achievable. Further, assume that there exists a single edge with capacity δ in the network. Then, there exists a network \mathcal{N}' with capacity $\mathcal{R}(\mathcal{N}') < \mathcal{R}(\mathcal{N}) - \delta$.

Proof: We show this by considering the example in Fig. 3, where two networks \mathcal{N}_1 and \mathcal{N}_2 are connected via a CC or a symmetrizable point-to-point AVC, resp., and an edge of capacity δ . Suppose that this connection also represents the min-cut of the overall network \mathcal{N} . For the AVC case, as the capacity of the symmetrizable AVC is either 0 or C_r , removing the δ -capacitated edge leads to a network capacity of $\mathcal{R}_{\text{AVC}}(\mathcal{N}') = 0$ according to Theorem 16. For the CC case the capacity of the CC is either \underline{C} or \bar{C} (see (45) and (46)). By removing the δ -capacitated feedback edge the network capacity is reduced from $\mathcal{R}_{\text{CC}}(\mathcal{N}) = \bar{C}$ to $\mathcal{R}_{\text{CC}}(\mathcal{N}') = \underline{C}$, where $\bar{C} - \underline{C}$ can be larger than δ . ■

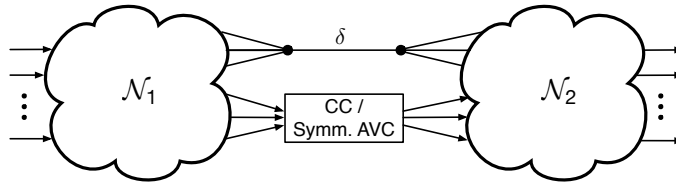


Fig. 3. The network \mathcal{N} consists of two arbitrary networks \mathcal{N}_1 and \mathcal{N}_2 connected by an edge with capacity $\delta > 0$ and a CC or alternatively, a symmetrizable AVC.

X. CONCLUSION

We have considered reliable communication over noisy network in the presence of active adversaries. This is modeled by a subset of independent point-to-point channels consisting of AVCs or CCs. For these cases we have identified scenarios for which the capacity of the corresponding noisy state-dependent network equals the capacity of another state-less network in which the AVCs or CCs are replaced by noiseless bit-pipes. Our results indicate that, in the network setting, the equivalent capacity of these channels is not necessarily equal to their capacity in an isolated point-to-point scenario. For example, the point-to-point AVC represents a pessimistic model for the action of an adversary, leading to zero capacity in some cases. We have shown that in a network setting such a pessimistic model becomes much more optimistic and leads to a positive rate if additional network connectivity exists between the head and the tail node of the AVC or CC under consideration. As most modern communication is performed in an underlying networking framework, this suggests that existing results may be insufficient for characterizing networks in the presence of active adversaries.

REFERENCES

- [1] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Medard, and M. Effros, “Resilient network coding in the presence of Byzantine adversaries,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, Jun. 2008.
- [2] S. Kim, T. Ho, M. Effros, and S. Avestimehr, “Network error correction with unequal link capacities,” in *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2009, pp. 1387–1394.
- [3] O. Kosut, L. Tong, and D. Tse, “Nonlinear network coding is necessary to combat general Byzantine attacks,” in *Proc. 47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2009, pp. 593–599.
- [4] —, “Polytope codes against adversaries in networks,” in *Proc. IEEE Int. Sympos. on Inform. Theory*, Austin, TX, Jun. 2010, pp. 2423–2427.
- [5] A. Lapidoth and P. Narayan, “Reliable communication under channel uncertainty,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [6] D. Blackwell, L. Breiman, and A. J. Thomasian, “The capacity of a class of channels,” *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.
- [7] J. Wolfowitz, *Coding Theorems of Information Theory*. Springer Verlag, 1978.
- [8] D. Blackwell, L. Breiman, and A. J. Thomasian, “The capacities of certain channel classes under random coding,” *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.
- [9] R. Ahlswede, “Elimination of correlation in random codes for arbitrarily varying channels,” *Probability Theory and Related Fields*, vol. 44, no. 2, pp. 159–175, 1978.
- [10] I. Csiszár and P. Narayan, “The capacity of the arbitrarily varying channel revisited: positivity, constraints,” *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, 1988.

- [11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.
- [12] M. Bakshi, M. Effros, and T. Ho, “On equivalence for networks of noisy channels under Byzantine attacks,” in *Proc. IEEE Int. Sympos. on Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 973–977.
- [13] R. Koetter, M. Effros, and M. Medard, “A theory of network equivalence—Part I: Point-to-point channels,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, 2011.
- [14] B. Hughes, “The smallest list for the arbitrarily varying channel,” *Information Theory, IEEE Transactions on*, vol. 43, no. 3, pp. 803–815, 1997.
- [15] T. Ho, M. Effros, and S. Jalali, “On equivalence between network topologies,” in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2010, pp. 391–398.
- [16] S. Jalali, M. Effros, and T. Ho, “On the impact of a single edge on the network coding capacity,” in *Proc. Information Theory and Applications Workshop*, San Diego, CA, Jan. 2011, pp. 1–5.
- [17] M. Langberg and M. Effros, “Network coding: Is zero error always possible?” in *Proc. 49th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2011, pp. 1478–1485.